

Sequential updating of a new dynamic pharmacokinetic model for caffeine in premature neonates

Short title: Updating of a pharmacokinetic model for caffeine

Sandrine Micallef¹, PhD., Billy Amzal², PhD, Véronique Bach³, PhD, Karen Chardon³, PhD, Pierre Tourneux⁴, MD, and Frédéric Y. Bois¹, PhD.

¹Institut National de l'Environnement Industriel et des Risques, Verneuil-en-Halatte, France.

²NOVARTIS PHARMA AG. Basel, Switzerland.

³Université Picardie Jules Verne Amiens, France

⁴Neonatal Unit and Intensive Care of the Pediatric Department of the Hospital of Amiens, France.

Acknowledgments

The authors wish to thank Dr. C. Diack for his helpful remarks and the anonymous referees for comments that improved our manuscript. Partial funding for this work was provided by the French Ministry of Research and Technology, project DIADEME, decision 02 C 0141. This work has also been supported by the research project BCRD-AP2001-DRC07 provided by the French Ministry of the Ecology and Sustainable Development and by the Regional Council of Picardie.

Corresponding author: Sandrine Micallef,
Experimental Toxicology Unit,
EA 3901 TOXI-DMAG,
Institut National de l'Environnement Industriel et des Risques,
Parc Technologique ALATA, BP 2,
60550 Verneuil en Halatte, France.
Tel: + 33 (0)3 44 55 61 68;
Fax: + 33 (0)3 44 55 66 05;
E-Mail: sandrine.micallef@ineris.fr

Category: Original Research Article.

Word count : 6391

Table of contents

Legends for figures.....	3
Abstract.....	5
Introduction.....	6
Methods	8
Pharmacokinetic Data.....	8
A Dynamic Model of Caffeine Biodistribution in Premature Neonates	9
One-Compartment Caffeine PK Model	9
Body Mass Model.....	10
Volume of Distribution Model	10
Clearance Model.....	11
Statistical Model.....	11
Measurement Error Model	11
Population Model	12
Bayesian Inference via MCMC.....	13
The Bayesian framework	13
Prior Distribution.....	13
Use of data for model calibration and adequacy checking	15
Model Calibration <i>via</i> MCMC	15
Sequential Updating.....	16
Results	18
Posterior Distributions	18
Model Fit Checking.....	19
Model Adequacy assessment	19
Comparison of Predictions from Previously Published Caffeine PK Models	20
Recalibration	21
Sequential Updating of Predictions Using Sequential Incoming Data.....	21
Discussion and Perspectives	24
Reference List	28
Tables.....	31

Legends for figures

Figure 1: Observed caffeine concentrations in blood as a function of postnatal age. Joined points indicate observations belonging to a same individual .

Figure 2: Directed acyclic graph of the hierarchical statistical model. Symbols are: P , prior distributions; μ_θ , population mean parameters; σ_θ , variance of population parameter; t , observation times; θ , unknown parameters; A_0 and M_0 are respectively duration of gestation and body mass at birth (both supposed exactly measured); X , the administered treatment; M_t , the measured body mass at time t ; cv_M , the variation coefficient of masses and Y_t , the measured caffeine concentration in blood. μ_{cvM} , and σ_{cvM} are respectively for mean and variance of cv_M parameters and σ_c for the variance of the experimental measurements. F is the dynamic bio-distribution model. Square nodes are for variables of known (or supposed known) values; circular nodes for unknown variables and the triangle represents a deterministic link.

Figure 3: Data versus model predictions. The left and the right graphs respectively concern body masses and caffeine concentrations in blood for individuals of the *training set*. On the right side graph, each line segment describes the 90%-credibility interval associated to the corresponding prediction. The 90%-credibility intervals have been plotted using σ_c value of the maximum posterior vector.

Figure 4: Caffeine concentration in blood (left) and body masses (right) as a function of time. Predictions have been calculated for one subject of the *training set* (Subject 9) and one subject of the *test set* (Subject 2). Means of the predictions are in bold lines, and the corresponding 90%-credibility intervals are plotted with thin lines. Data are represented by dots. Note that 90%-credibility interval of the predicted masses is not represented for the subject of the *training set* (subject 9) as it is very thin.

Figure 5 : Relative residuals for caffeine concentration as a function of postnatal age. Four different models are represented: Thomson *et al.* ^[1] (triangles), Lee *et al.* ^[2] (squares), Falcao *et al.* ^[3] (empty circles) and the one we propose (plain circles). Relative residuals correspond to the relative differences between *test set* data points and mean caffeine concentration predictions obtained by Monte Carlo simulations.

Figure 6: Body mass predictions as a function of the time for patient 31. The 5%, and 95% quantiles of predictions are plotted using gray lines, the mean of predictions using dark bold lines. Data are represented by the dots. Panel A shows the predicted masses calculated from the original posterior-sampled parameters, using the covariates of the patient at birth, only. Predictions of panel B use, in addition, body mass data via the updated posterior parameters obtained by a particle algorithm.

Figure 7: Caffeine concentration predictions as a function of time for patient 31. The 5% and 95% quantiles of predictions are plotted using thin lines, the mean of predictions using bold lines. The data point is represented by dot. Panel A shows the predicted caffeine concentration in blood calculated from original posterior sampled parameters and using only the covariates of the patient. Predictions of panel B and C are performed using the posterior parameters updated *via* a particle algorithm. In panel B, only data masses are used for parameter updating while only measured caffeine concentration is used to update predictions in panel C.

Figure 8: Histograms of the marginal posterior distributions of the population elimination model parameters. Top row graphs give the marginal distributions of the *original* posterior. Bottom row darker histograms give the marginal updated distributions, using one caffeine concentration data point for subject 31. Bottom row unfilled histograms recall the *original* posterior.

Abstract

Background: Caffeine treatment is widely used in nursing care to reduce the risk of apnea of premature neonates. To check the therapeutic efficiency of the treatment against apnea, caffeine concentration in blood is an important indicator. The present study aims at building a pharmacokinetic model as a basis for a medical decision support tool. **Methods:** A new dynamic pharmacokinetic model is proposed to predict caffeine disposition in premature neonates. Time dependence of physiological parameters is introduced to describe rapid growth of neonates. To take into account the large variability in the population, the pharmacokinetic model is embedded in a population structure. The whole model is inferred within a Bayesian framework. To update caffeine concentration predictions as data of an incoming patient are collected, we propose a fast method such it can be used in a medical context. This involves the sequential updating of model parameters (at individual and population levels) *via* a stochastic particle algorithm. **Results:** Our model provides better predictions than the ones obtained with models previously published. We show through an example that sequential updating improves predictions of caffeine concentration in blood (reduce bias and length of credibility intervals.) The update of the pharmacokinetic model using mass and caffeine concentration data is studied. It shows how informative caffeine concentration data are in contrast to mass data. **Conclusion:** This study provides the methodological basis to anticipate caffeine concentration in blood, after a given treatment if data are collected on the treated neonate or not.

Keywords: Pharmacokinetic compartmental model, population model, Bayesian inference, MCMC algorithms, particle algorithms, caffeine, neonates

Introduction

Premature neonates suffer from various physiological deficiencies due to their immaturity at birth. In particular, they can suddenly stop breathing (apnea) spontaneously and quite frequently^[4]. To reduce the occurrence of apneas, a caffeine treatment is often administered to premature neonates^[5-7]. This treatment may last several weeks and can induce various secondary effects (on ventilation, sleep, metabolism, weight, glomerular filtration, *etc.*)^[8-12]. Unfortunately, pediatricians have little knowledge about the actual occurrence of these effects, which appears to be widely variable from an individual to another. To balance treatment benefits and adverse effects, pediatricians try to keep caffeine blood concentration within a "therapeutic" interval. This is checked *via* blood sampling. Since sampling blood causes trauma on such fragile infants, the number of blood samples is limited. Therefore, it is not possible to collect extensive information on caffeine concentration time course in a given patient and pediatricians refer to a standard protocol of administration^[13, 14]. In doing so, they control treatments or times to sample "on average" in the patients' population. The aim of this paper is to define a PK model to be used as a basis for a medical decision tool to help pediatricians in individualizing caffeine administration. The developed PK model allows the predictions of caffeine time-course in a new patient for whom only easily collected covariates (duration of gestation and body mass at birth), and possibly, a few caffeine concentration data would be available. For this, we need to take advantage of all available information: data extracted from medical files, pediatricians' experience and published data on neonate physiology and caffeine population pharmacokinetics^[1-3]. In this context, a Bayesian approach is efficient, as it allows an easy aggregation of such widely different and uncertain information sources^[15-17] and naturally allows sequential learning.

To describe caffeine concentration in blood, a compartmental PK model which is an extension of the one proposed by Amzal *et al.* ^[18], is proposed. As caffeine is hydrophilic and distributes in body water^[19], our pharmacokinetic model has only one compartment. Its main parameters are the volume of distribution and the caffeine clearance. Moreover, as volume of distribution (depending on the amount of total body water) and metabolism change significantly during the few weeks of the treatment^[20-22], we propose a dynamic model with time-dependent parameters. To take into account inter-individual variability, a hierarchical population model is defined and calibrated using historical data.

As caffeine elimination is widely variable in infant populations^[1-3], a more efficient PK model need to integrate new information on the treated patient. We show here how caffeine concentration predictions of the treated patient can be updated by running a particle algorithm using recorded body mass and/or caffeine concentration data. The proposed algorithm has the advantage to update both individual and population parameters on the basis of new individual data. Therefore there is no need to re-perform time-consuming calibration process to estimate new population parameters when incorporating this new information.

To our knowledge, any tool for caffeine dose individualization is not used in pediatric services, until now. Even if methods for dose individualization have been proposed in the literature for other drugs^[23-25] none has been applied to caffeine dosing. We show here how a population pharmacokinetic model can be calibrated using a particle algorithm in order to guide individual caffeine dosing in a reasonable time. Formal optimization of the treatment is not treated; we refer to Amzal *et al.* ^[18] and references therein for details on that question.

In the present study, we start by the modeling of the caffeine bio-distribution, explaining how physiological parameters evolve with time. The following section is devoted to defining the

statistical model, to quantify sources of uncertainty. It also describes the Bayesian framework in which the inference problem is casted, recalling its principle and defining the chosen prior distributions. The next section details the model checking and model adequacy procedures used. We demonstrate, in the last section, how the model predictions can be sequentially updated with incoming data on new treated patients, and how this can be used for designing a medical decision support software. Finally, this approach and its results are discussed and put in perspective.

Methods

Pharmacokinetic Data

The pharmacokinetic data come from medical files of 35 neonates treated in the Neonatal Unit and Intensive Care of the Pediatric Department of the Hospital of Amiens (France). These files are the only available in the set of medical files of the neonates admitted in the nursing care department during years 1998, 1999 and 2003. Files were pre-selected such they meet the following requirements:

- contain at least one caffeine concentration measurement,
- the treatment lasts 3 days at least,
- the treatment goes at least until the post conceptional age of the neonates is 35 weeks.

The typical treatment administered to neonates consists of a charging injected dose of 20 mg/kg possibly spread out on the first two days, then a regular oral dose of 5 m/kg every day from the third day. The treatment usually lasts until the post conceptional age of the neonates reaches 35 weeks. Naturally, these doses could have been modified by the pediatricians for clinical reasons. Observed concentrations of caffeine in blood were collected as requested by the physician. It does not exist any sampling scheme (see Figure 1).

[Figure 1 about here]

For each patient, the extracted data Y_t consisted of one to four concentrations of caffeine measured in blood after specific treatment X (timing, injected concentration or quantity *given per os*), and measured daily body masses, M_t . An example of collected data (for two patients) is given in Figure 4; the observed caffeine concentrations in blood for all patients are represented on Figure 1. Covariates such as duration of gestation A_0 and body mass at birth M_0 were also recorded. These data are available upon request.

The average of the body masses at birth for our neonates population ($N=35$) is 1.34 kg with a standard deviation of 0.35 kg. The smallest recorded masse at birth is 0.660 kg while the largest is 2.170 kg. The duration of gestation of these premature neonates is 29.1 weeks long on average (with a standard deviation of 2.1 weeks and a range of [23, 32] weeks of gestation).

A Dynamic Model of Caffeine Biodistribution in Premature Neonates

One-Compartment Caffeine PK Model

According to published pharmacokinetic studies^[1-3], caffeine distribution in the body is well described by a one-compartment model^[19]. The body is therefore modeled by one homogeneous compartment of volume $V(t)$ at time t . Note that the time origin ($t=0$) corresponds here to birth time. Caffeine is eliminated from the central compartment with clearance $CL(t)$. The time-course of caffeine concentration $c(t)$ (in mg/L) is therefore described using a first-order elimination. Input drug is modeled as an IV bolus using a zero-order absorption even in case of oral dosing (the assumption here is that absorption is very rapid). This assumption is made because necessary data to identify absorption rate are incomplete.

Body Mass Model

From simple examination of the body mass data, it appears that the growth of premature neonates can be, in a first approximation, divided into two distinct periods. The body mass decreases during the first period of length τ , while it subsequently increases during the second period. We propose linear relationships to describe that variation during the two periods. These functions depend on the covariates M_0 (body mass at birth), A_0 (post-conceptional age at birth supposed exactly known) and are specific to each individual:

$$\begin{aligned} M(t) &= M_0 + \alpha_1 t \quad \text{if } t \leq \tau \\ M(t) &= M_0 + \alpha_2 (t - \tau) \quad \text{if } t > \tau \end{aligned}$$

where α_1 and α_2 are the slopes of the straight line respectively decreasing and increasing of each phases. M_τ corresponds to the body mass at τ ($M_\tau = M_0 + \alpha_1 \tau$).

Volume of Distribution Model

The volume of distribution is assumed to vary linearly with body mass:

$$\begin{aligned} V(t) &= M(t) f_{25} + \beta (A_0 + t - T_{25}) \quad \text{if } t \leq \tau \\ V(t) &= M(t) f_\tau \quad \text{if } t > \tau \end{aligned}$$

where f_{25} is defined as the ratio of the distribution volume over body mass at post-conceptional age 25 weeks (time T_{25}) and f_τ is the value of that ratio at time τ ($f_\tau = \frac{V(\tau)}{M_\tau}$). β is the slope of the

variation of the volume-over-mass ratio as a function of time. It is defined as follows:

$$\beta = M_0 \frac{f_\tau - f_{25}}{A_0 + \tau - T_{25}}$$

The slope β is set such as the ratio volume over mass at post-conceptional age 25 weeks is f_{25} and as the ratio volume over mass at τ is f_τ . As the degree of plasma protein binding is quite low for caffeine^[26], the

volume of distribution is assumed comparable to total body water. In this case, f_{25} and f_τ are approximately equal to fraction of water at time T_{25} and τ .

Clearance Model

Several authors showed an increase in elimination of caffeine with postnatal age^[21, 27, 28]. According to all of them, caffeine elimination increases as a function of time and reaches a plateau after 3 to 4.5 months. Taking into account the duration of treatment (less than two months in general), we only modeled the first period and assumed that caffeine clearance increases linearly with time with a slope parameter γ and a constant at origin ($t=0$) CL_0 :

$$CL(t) = CL_0 + \gamma t.$$

Statistical Model

The statistical model quantifies both uncertainties (on data measurements as well as modeling) and sources of variability. Data measurements as well as modeling and intra-individual variability errors are taken into account by using a measurement error model. To model uncertainty on parameters and their variability, probability distributions are used. Inter-individual variability is quantified in the framework of a population model^[29, 30] by estimating its population parameters. Given the above bio-distribution model, the caffeine concentration in blood of a given patient is conditioned by the covariates M_0 and A_0 and the parameter vector $\theta = (\alpha_1, \alpha_2, \tau, f_\tau, f_{25}, CL_0, \gamma)$.

Measurement Error Model

The caffeine concentration in blood actually observed is affected by measurement errors which are assumed to be independent and log-normally distributed with a geometric mean of one and a geometric standard deviation σ_c . Likelihood is then given by:

$$\log(Y) \sim N(\log(F(X, \theta)), \sigma_c^2)$$

where the function F corresponds to the PK model described above.

A normal distribution (with zero mean and standard deviation σ_M) measurement error model is also specified for body masses M_i :

$$M(t) \sim N(M(t), \sigma_M^2)$$

To take into account inter-individual variability of body masses (due, for example, to urinary excretion), the standard deviation of measured body masses has been considered as individual-specific. It has been chosen as depending on the body mass of the neonate and on the individual parameter cv_M :

$$\sigma_M = cv_M \times M_i$$

Population Model

The pharmacokinetic model has been embedded in a population model which considers that each subject parameter values θ are drawn from a statistical distribution G with given population mean μ_θ and variance σ_θ (Figure 2).

[Figure 2 about here]

The proposed population model has two major components: the individual and the population levels. At the individual level, parameters θ are assumed to be normally or log-normally distributed (see Table I) with population means μ_θ and standard deviations σ_θ . Truncation bounds of the distributions are set on the basis of the limits for plausible values. For each subject, the expected values of body mass and caffeine concentration in blood are a function F of administration X , depending on parameter θ , and the individual's covariates. At the population level, cv_M is assumed to be distributed according to a truncated normal distribution with mean μ_{cvM} and standard deviation σ_{cvM} (Table I).

[Table I about here]

Bayesian Inference via MCMC

The Bayesian framework

The population model described above is fitted to observed data, Y_t , with Bayesian techniques^[16, 17]. The Bayesian approach yields a sample θ of parameter values from their joint posterior distribution. From Bayes' theorem, the joint posterior distribution of parameters is proportional to the prior distributions of parameters multiplied by the data likelihood. The posterior is then an update, using collected data, of what it is known about parameters prior to the experiment.

Prior Distribution

The first step of the Bayesian approach consists in defining prior distributions that quantify the information coming from pediatricians expertise or from the scientific literature. Expertise of clinicians is then entered in the model through its structure and through the use of prior distributions parameters. Prior distributions on population parameters and their associated parameters are summarized in Table I. On the basis of available information, we specify a prior that covers the range of value that is deemed reasonable for each parameter. When we deal with mean parameters, we use truncated normal or uniform distributions as prior while for standard deviation parameters truncated lognormal or loguniform laws are used. For standard deviations, lognormal prior distributions are chosen when little prior knowledge about the parameter is available. For parameters on which no prior information is found, loguniform distribution is used. This distribution is preferred to the more standard Inverse Gamma distribution since it does not require to define mean and variance (as Inverse Gamma does) and only uses truncation bound.

For parameters related to body mass (α_1 , α_2 , τ), priors are elicited^[31-33] by encoding the pediatricians' knowledge. From their experience, neonates loose, on average about 25g/day during between the 3 to 20 first days, then gain about 25g/day. This is used to build priors on $\mu_{\alpha_1}, \mu_{\alpha_2}, \mu_{\tau}$. Information on $\sigma_{\alpha_1}, \sigma_{\alpha_2}, \sigma_{\tau}$, are less easily elicited and rather flat priors are used.

Since the degree of plasma protein binding is quite low for caffeine, we assume that the ratio volume over mass is approximately equal to fraction of body water. Thus, for f_{25} and f_{τ} , priors are proposed after referring to the percentage of water in body reported by the International Commission on Radiological Protection^[20]. For parameters related to clearance, CL_0 and γ , prior distributions are defined after re-parameterization of values from the literature^[1-3]. Globally, all the previous studies agree and give the same magnitude order for the mean of the clearance at birth, μ_{CL_0} (estimated between 0.006 and 0.03 L/h with an estimated standard deviation around 0.003) and for the shape of temporal evolution of clearance μ_{γ} (estimated between $1e^{-6}$ and $5e^{-5}$ with an estimated standard deviation around $3e^{-6}$). We have chosen mean of μ_{CL_0} and μ_{γ} such they are a compromise between the published studies. To be not too restrictive and to take into account possible differences between studied populations, standard deviations are chosen a bit higher than those estimated in the literature. Population variance of clearance parameters ($\sigma_{CL_0}^2$ and σ_{γ}^2) are assigned prior distributions truncated between 1.1 and 3 corresponding to a population variability comprised between 10 and 300%: less than 10 % of variability for a clearance parameter in the population seems incredible and 300% seems large enough since inter-individual variability of caffeine clearance was estimated to be around 80% for adults^[34] and is likely a bit larger for neonates.

Use of data for model calibration and adequacy checking

The data are used to calibrate (fit), check and assess adequacy of the model. The adopted approach for model adequacy checking is *cross-validation* ^[35, 36]. To that effect, data files are divided into two groups: the *training set* and the *test set*.

We start with the *training set* for model calibration (estimation of the different parameters) and checking the fits. Then, next we use the calibrated model to check its adequacy with respect to the *test set*. We also perform another evaluation of the model which consists in calculating the percentages of caffeine concentration and body mass data points from the *test set* falling within different credibility intervals for predictions. The results of these tests are discussed further below.

Both sets of data are formed by a random binomial assignment with parameters equal to ($p=1/3$, $N=35$). Twenty-four medical files are drawn for the training set (calibration and checking the fits) and the other eleven files are assigned to model adequacy checking (test set).

This process allows us to validate the model. Once this is done, we recalibrate the model using the full data set (35 subjects) to take into account all available information in our prediction procedure. This recalibration step gave what we call in the sequel, *original posterior distribution*.

Model Calibration *via* MCMC

We assessed the model parameters by sampling their values from the joint posterior distributions of all population and individual parameters, using the data from the 24 *training set* subjects. For this, we used Markov Chains Monte-Carlo simulations^[37, 38] (Metropolis-Hastings sampler) performed by MCSim software^[39] version 5.0.0. After 40,000 iterations of 3 parallel Markov chains, their convergence is checked by calculating the Gelman and Rubin \hat{R} ratio on the 20,000 last iterations^[16]. The highest \hat{R} computed is 1.12, showing that all chains had approximately

converged. To get a sample from the targeted posterior distribution, we took one in 10 vector out additional 20,000 iterations of each chain after convergence checking (iteration 40,000 to 60,000) leading to a posterior sample of 6,000 parameter vectors. Running the three MCMC chains took about one day (on a pentium 4, microprocessor PC 2.8GHz).

Sequential Updating

For an incoming subject (just after birth), parameters can be guessed on the basis of population distribution (i.e., by sampling from those distributions). However, when data are collected on a subject during treatment, the parameters for that individual can be updated. For parameter updating, it is possible to re-perform the calibration process described above using all available data including the newly recorded ones. This involves re-running Markov chains. But this procedure is not applicable in practice by pediatricians because it is time-consuming. Thus, we propose to apply a particle algorithm^[40-43] to sequentially update the posterior (individual and population) parameter sample obtained by calibration (*original posterior*). Since they start from the original posterior rather than from the priors of Table I, calculations converge faster than the *classical* calibration procedure^[44]. Individual parameters and population parameters can be jointly and quickly updated. Other alternative methods to standard MCMC calibration have been proposed (MAP Bayesian, Multiple Model, and Interacting Multiple Model)^[29, 45] to update predictions of the new patient. Our method generalizes MAP approach (consider the whole distribution instead of maximum *a posteriori*) and allows the updating of population parameters which was not proposed with MM or IMM methods, to our knowledge. The updating of population parameters would allow predictions on another future patient using the information brought by the incoming one.

The proposed updating method assumes that a current sample is available and consists in few steps. Call $(\theta_1, \dots, \theta_N)$ our *original posterior* sample of parameters, which approximate a distribution $p(\theta)$, and let's denote y the new coming data. Each θ_i is also called a "particle" in this framework. We want to transform our sample into an another one (of same size N) from $p(\theta|y)$.

The detailed updating algorithm can be described as follows:

- For each particle $i=1..N$, compute its normalized weight w_i (as in the Sampling Importance Resampling scheme of Rubin ^[46, 47]

$$w_i \propto \frac{p(\theta_i|y)}{p(\theta_i)} \propto p(y|\theta_i)$$

These weights are easy to compute since they take the form of likelihood terms normalized by the sum of all likelihood.

- Resample N particles $(\theta'_1, \dots, \theta'_N)$ from $(\theta_1, \dots, \theta_N)$ according to the multinomial distribution $M(N, w_1, \dots, w_N)$
- To avoid degeneracy problems, enrich the sample using a Markov kernel Q whose stationary distribution is $p(\theta|y)$. For each $i=1..N$, we draw $(\theta''_1, \dots, \theta''_N)$ from $(\theta'_1, \dots, \theta'_N)$. By definition, this transforms an approximated sample $(\theta'_1, \dots, \theta'_N)$ from $p(\theta|y)$ into another sample $(\theta''_1, \dots, \theta''_N)$ from $p(\theta|y)$, but now richer (more diverse). This final sample is the updated parameter sample.

A convenient choice for the Markov step Q is a Metropolis-Hastings move as used in standard MCMC ^[37, 38]. We chose a Metropolis-Hasting step with an independent sampler ^[48] in which candidates for population parameters are drawn from the *original posterior* sample and individual parameter from population distribution. To perform updating procedure, we use Matlab software.

Results

Posterior Distributions

The MCMC-calibrated model has been first checked by analyzing the initial posterior vectors' sample (data not shown, see Table II for a summary of final posterior estimate).

Means of mass model parameters μ_{a1} and μ_{a2} were quite well appraised *a priori* by the experts, but calibration has decreased the standard deviation of these two parameters. Premature neonates lose, on average, 30 grams per day until the end of the sixth day, then gain, in a second period, 23 grams per day on average. Body mass loss and the length of time during which decrease occurs are quite variable between individual (around 40% for both). The values of body weight gain are less variable in this population (around 25%).

Posterior means of the population parameters related to the volume of distribution (μ_{f25} and $\mu_{f\tau}$) are close to the priors. This is partly due to the very informative nature of the prior distribution for these two parameters. The mean of the volume-over-mass ratio in population is equal to 85% at post-conceptual age 25 weeks and 75% at time τ . In parallel, calibration has increased knowledge on population variability (The CVs of σ_{f25} and $\sigma_{f\tau}$ decreased from 20% to 16.3% and 15.3% respectively).

On average, clearance at birth is about 0.3 L/day. For our neonates (mean body mass: 1.340 kg and mean distribution volume: 85% of mass), the half-life of caffeine is $t_{1/2}=2.6$ days. Clearance at birth, CL_0 , and slope of the temporal linear evolution of clearance, γ , are quite variable in this population (CV of CL_0 about 27% and CV of γ equal to 77%).

Model Fit Checking

Figure 3 shows model-predicted values using the joint maximum posterior parameter vector *versus* their observed counterparts.

[Figure 3 about here]

The posterior sample gives quite a good fit to the body mass observations. The average of the relative errors for mass is about 0.1% with a minimum around -12.4% and a maximum of 14%. Predicted caffeine concentrations in blood are not that good; the relative error is 8.5% in average with a minimum and a maximum respectively equal to -108% and 48%. Figure 4 shows the average of predicted caffeine concentrations and body mass predictions over time, for one subject of the *training set* (Subject 9). As already observed on Figure 4, model predictions well describe the data.

[Figure 4 about here]

Model Adequacy assessment

Using the posterior sample, for each patient of the *test set*, mean curve and 90%-credibility intervals for predictions of body mass and caffeine concentration have been constructed (from Monte Carlo simulations) as a function of time (Subject 2 is the only plotted on Figure 4). For all the subjects of the *test set*, most of the caffeine concentration data fall in the credibility interval and body masses data are well estimated.

The empirical percentages of data points falling within different credibility intervals are close to the credibility percent coverage (last line of Table III), showing that the shapes of the statistical distributions chosen to model measurement errors are appropriate. The fact that observed percentages for caffeine concentration tend to be lower than the theoretical ones in the tails could

be explained by the small number of caffeine concentration data points available in the *test set* (N=21 for 11 subjects).

Comparison of Predictions from Previously Published Caffeine PK Models

Three population PK models for caffeine have been previously proposed ^[1-3]. Caffeine concentration predictions for data belonging to the *test set* are made using these three models. Two of these models (Lee's and Falcao's) cannot predict body masses of neonates. Using set body masses (as observed at birth, for example) would be very penalizing for these models, so we choose to use the actually measured body masses, which is to their advantage. Monte Carlo simulations are run on the basis of the published parameter distributions to obtain credibility intervals on caffeine concentration predictions. For each model previously proposed average mean square errors (AMSE) and percentages of adequacy data points falling in different "theoretical" credibility intervals are given in Table II.

[Table II about here]

Our model better assesses uncertainty on caffeine concentration predictions: observed percentages of data points falling in the "theoretical" credibility intervals are better than those obtained with previous models. These models have higher bias or over-dispersion (higher AMSE)

than our model. Relative residuals ($\frac{Y_{obs} - \bar{Y}}{Y_{obs}}$, Y_{obs} observations and \bar{Y} the estimate) of each model

are plotted against postnatal age on Figure 5. It appears that our model gives a systematic prediction error at early time point (between the third and the eighth day of treatment). This point is discussed further below. However, most of the times, the proposed model gives better mean predictions than the other models for this data set.

[Figure 5 about here]

Recalibration

Again, three chains of 60, 000 iterations are run. One out of 10 of the last 40, 000 simulations of the three chains are recorded, yielding 12, 000 parameter sets. The Gelman-Rubin convergence diagnostic is less than 1.06 for all parameters. Using this very final sample, summary statistics for marginal posterior distributions of the population parameters are given in Table III.

[Table III about here]

Compared to posteriors obtained during the first calibration step (involving 24 subjects), the estimates have not drastically changed. Estimated statistics (means and standard deviations) of population parameters have changed by less than 33%. Parameters related to α_1 and CL_0 are the most significantly updated. μ_{α_1} is higher than previously; σ_{α_1} , μ_{CL_0} and σ_{CL_0} have also increased. The estimated uncertainty of these parameters is slightly higher than after the initial calibration step. This is due to fact that the estimates of the individual parameters CL_0 and α_1 for the subjects of the *test set* are not fully centered in the previous population distribution.

Sequential Updating of Predictions Using Sequential Incoming Data

Obviously, more accurate predictions and credibility intervals would arise from taking into account the incoming data (body masses, blood samples, *etc.*) about subjects as they are clinically followed. This sequential update can easily be performed by correcting the previous posterior distributions using the new data. Since this update would not induce a drastic change, it is not necessary to run new MCMC chains. We demonstrate here the use of an efficient way to do it *via* particle methods^[40].

Let us illustrate the method through an example. Sequential updating is performed for patient 31 (of the *test set*). We examine how different kinds of data can improve caffeine concentration predictions. The update is first done using mass data, then using caffeine concentration data. The

update could also be done with both kinds of data (mass and caffeine concentration data); the procedure would be same. Performing update with body mass data and caffeine concentration data separately allows to illustrate how informative caffeine concentration data are in comparison with body mass data.

Neonate 31 is born at $A_0=30$ weeks, with body mass $M_0=1.575\text{kg}$. The daily body masses were measured during the 9 first days of treatment (Figure 6). The update is implemented first on the basis of recorded mass, using an original posterior sample of size of 12,000. An updated parameters sample is then obtained.

The updated posterior distributions for α_1 , α_1 and τ are more accurate than the initial guess. Body masses dramatically increase knowledge about the time evolution pattern of mass for the patient. Figure 6 shows, for that subject, the predicted masses generated from the *original posterior* sampled parameters (panel A) and the predictions generated from updated parameters (panel B). Masses are clearly much better predicted in panel B. This updating procedure took about sixty seconds (on a pentium 4, microprocessor PC 2.8GHz running Windows XP), whereas running the three chains of the calibration procedures took about one day. The proposed updating procedure is about 1500 time faster than MCMC initial calibration.

[Figure 6 about here]

It is interesting to see how update can improve caffeine concentration predictions. Caffeine concentration in blood was measured for this subject on the third day of treatment ($c=5.2$, mg/L). The predictions derived from the *original posterior* distribution are quite bad (Figure 7A) and the predictions obtained from updated sample using the only body mass data are not good either (Figure 7B). Whereas predictions for masses are much better after mass parameter updating (Figure 6), caffeine concentration predictions on Figure 7B seems unchanged (the distribution tails below 5%tile and above the 95%tile do change in fact).

The same updating method is then performed using only the caffeine concentration observation (Figure 7C). The concentration data now falls within the 90%-credibility interval and the mean estimate (9.7 mg/L) is also nearer the data after updating than before (17.6 mg/L). The concentration data is very informative relatively compared to the *original posterior*, especially on clearance, as expected. The gain on precision of predictions brought by the updating using caffeine concentration data can be measured by comparing the spread of the 90%-credibility interval on caffeine concentration predictions before and after the update. The mean range between the lower and the upper bounds of the credibility interval (respectively corresponding to the 5th and the 95th %tiles) is equal to 26.2 mg/L before updating, 25.9 mg/L after updating using mass data and has been reduced to 13.2 mg/L after updating using caffeine concentration data.

[Figure 7 about here]

CL_0 marginal updated distribution for individual'31 has been transformed (in shape and mean) after updating (the mean of CL_0 has changed from 0.017 to 0.033 and the standard deviation has staid around 0.008). On the other hand, γ distribution has not been much revised (the average of γ has changed from 1.043×10^{-5} to 1.087×10^{-5} and the standard deviation from 0.889×10^{-5} to 1.020×10^{-5}).

An important concern is to decide when to update the distributions. It would not be optimal to do it every day, for example after each body weight measurement, because of the low sensitivity of caffeine elimination parameters to those measurements. A pragmatic general answer can be found in Chopin 2002 ^[40]. In our case, a reasonable choice might be to perform the updating as soon as the observed body mass goes out of the predictions' (90%) credibility interval. Moreover, since the caffeine concentration data are strongly informative, it seems that update should be done after each new caffeine concentration measurement.

The updating procedure improves predictions for an individual as new data are collected, and thus, allows to guide dosing for this particular patient taking into account all its physiological characteristics (given by the covariates and the data). At the same time, this procedure also updates the joint population parameters distributions since MCMC calibration produces samples from the overall joint distribution of population and individual parameters. Figure 8 shows histograms of the marginal updated distributions for population mean and standard deviation of CL_0 and γ . As only one caffeine concentration data point on one subject has been used to update these population parameters, their distributions have not drastically changed. Nevertheless, μ_{CL_0} and σ_{CL_0} distributions have been slightly translated to the right after updating, corresponding to greater values. This is coherent with the high value of CL_0 for that individual. On the other hand, σ_γ distributions before and after updating shows that σ_γ has been slightly reduced because the estimated γ for the new subject is well centered with the original population distribution.

[Figure 8 about here]

Discussion and Perspectives

The present study aims at providing the methodological basis for a quantitative decision support software for pediatricians who would like to anticipate neonates' caffeine concentration in blood, for a given treatment. Fitting a population model to individual data allowed us to make predictions for new subjects, given their covariates at birth, and a sample of random vectors of parameters from the posterior population distribution. These simulations give an "average prediction" for future caffeine concentrations during treatment, as well as credibility intervals for them. We present a fast and easy way to update those predictions on the basis of subject specific data on body mass and caffeine concentration in blood.

Three caffeine distribution population models have already been proposed^[1-3]. Thomson *et al.*'s model assumes that clearance depends on body mass at birth and does not account for mass changes over time. Similarly, caffeine distribution volume is also assumed to stay constant during treatment period. Making predictions over weeks, in such a context, can be misleading since the physiological variable of the premature neonates change significantly during the first trimester of life. Falcao *et al.* and Lee *et al.*'s models consider parameters (clearance and volume) as dynamic but requires infant body mass values at the time at which caffeine concentration is to be predicted. Lacking a body mass model, such models cannot be used for predictions over several days or weeks. However, even with a perfect mass model (i.e. using actually measured body masses), these two models do not give predictions as good as those of our new model (see Figure 5). Overall, the model we propose gives the best predictions for our neonate population. Two points need however to be stressed:

- Its mean predictions are still quite far from some data points (7 over 21 relative residuals errors are over 50%) revealing that the model might still be improved. Moreover, relative residual errors show a trend with post-natal age. Predictions of all four models over estimates just after birth and underestimates after about 25 days of postnatal age (Figure 5). This implies the models are not able to capture all information contained in the data.
- Our model works quite well with local (Amiens) neonates. We do not know if it would be the best at describing Thomson's, Lee's or Falcao's data. We would recommend for a new calibration when going to new populations.

The population variability in caffeine clearance is relatively large and is the cause of the relatively large width of credibility intervals for the caffeine concentration predictions. This actually reflects the daily reality met by physicians. Part of this variability might be explained by introducing covariates which have not been taken into account here (as caffeine or tobacco

maternal consumption during pregnancy, other administered treatment than caffeine, for example).

Another improvement could concern the model itself, and more precisely the absorption model. We made the assumption that oral absorption is very rapid. This is questionable. Not describing precisely the absorption process could generate errors in predictions at early time points. Given that misspecification the model could over-estimate the speed of caffeine bioavailability after the third day (beginning of *per os* administration). This over-estimation could lead to an under-dosing over that period. A first-order oral absorption model might be better in this situation. However, after about the eighth day, caffeine absorption by oral route seems more or less quite well represented with a zero-order absorption rate (see Figure 5). We also made some choices which could be reconsidered, such as the linear form for the temporal evolution of physiological parameters, or the relative error to model the intra-individual variability in the body mass. Non-parametric modeling^[49, 50] of population distribution could also be an interesting approach. Non-parametric models would relax hypotheses on parameter distributions. They could allow identification, for example, of bimodal population distributions, which could help in determining new explanatory covariates.

As different updating methods exist, it could be interesting to compare the different approaches of sequential updating (MAP Bayesian, Multiple Model, and Interacting Multiple Model). It would also be interesting to compare our method to other ones^[23-25] already proposed for dose individualization by applying them to caffeine dosing, for example.

A therapeutic dosing support tool should incorporate a sequential updating procedure such as the one detailed before. Admittedly, our procedure needs to be tuned (jump function, number of particle)^[40] but it is fast, automatic and can be programmed easily in a medical support tool (currently under development). The only requirement for the user is to enter the newly collected

data. The pediatrician does not need to know details of modeling or statistics. The therapeutic support tool would, then, give individualized predictions on the basis of information collected on the treated patient. The updating using caffeine concentration data can give better estimate (less biased) and as the same time, reduce the spread of credibility interval of caffeine concentration predictions (by around a factor 2, in our example). The proposed particle algorithm allows to quickly update individual parameters distribution. Moreover, population parameters distributions are updated simultaneously so that all available information from past or current patients will be available for any new subject entering the hospital.

To validate the benefit of the use of such a pharmacokinetic tool, a clinical trial could be based on a judgement criteria including observed effects after the comparison of two patient groups: one receiving the usual dosing and another one for which the dosing is assisted by the software tool.

Finally, this technique was applied to caffeine as a demonstration study but could also be applied to other drugs with a narrower therapeutic window. Indeed, that would be all the more useful. Drugs for which the target serum concentration is close to the concentration limit leading to severe adverse effects, would certainly benefit to this method since it allows a quick and individualized estimation of the drug concentration using a limited number of measurements.

Reference List

1. Thomson AH, Kerr S, Wright S. Population pharmacokinetics of caffeine in neonates and young infants. *Ther Drug Monit* 1996;18(3):245-53.
2. Lee TC, Charles B, Steer P, Flenady V, Shearman A. Population pharmacokinetics of intravenous caffeine in neonates with apnea of prematurity. *Clin Pharmacol Ther* 1997;61(6):628-40.
3. Falcao AC, Fernandez de Gatta MM, Delgado Iribarnegaray MF, Santos Buelga D, Garcia MJ, Dominguez-Gil A, et al. Population pharmacokinetics of caffeine in premature neonates. *Eur J Clin Pharmacol* 1997;52(3):211-7.
4. Daily WJ, Klaus M, Meyer HB. Apnea in premature infants: monitoring, incidence, heart rate changes, and an effect of environmental temperature. *Pediatrics* 1969;43(4):510-8.
5. Aranda JV, Turmen T, Davis J, Trippenbach T, Grondin D, Zinman R, et al. Effect of caffeine on control of breathing in infantile apnea. *J Pediatr* 1983;103(6):975-8.
6. Aranda JV, Gorman W, Bergsteinsson H, Gunn T. Efficacy of caffeine in treatment of apnea in the low-birth-weight infant. *J Pediatr* 1977;90(3):467-72.
7. Aranda JV, Grondin D, Sasyniuk BI. Pharmacologic considerations in the therapy of neonatal apnea. *Pediatr Clin North Am* 1981;28(1):113-33.
8. Zanardo V, Dani C, Trevisanuto D, Meneghetti S, Guglielmi A, Zacchello G, et al. Methylxanthines increase renal calcium excretion in preterm infants. *Biol. Neonate* 1995;68(3):169-174.
9. Gillot I, Gouyon JB, Guignard JP. Renal effects of caffeine in preterm infants. *Biol. Neonate* 1990;58:133-136.
10. D'Urzo AD, Jhirad R, Jenne H, Avendano MA, Rubinstein I, D'Costa M, et al. Effect of Caffeine on ventilatory responses to hypercapnia, hypoxia, and exercise in humans. *J. Appl. Physiol.* 1990;68(1):322-328.
11. Aranda JV, Clozel M. Pharmacologic effects of caffeine and theophylline in the premature infant. *Dev Pharmacol Ther* 1982;4 Suppl:165-72.
12. Karacan I, Thornby JI, Anch M, Booth GH, Williams RL, Salis PJ. Dose-related sleep disturbances induced by coffee and caffeine. *Clin. Pharmacol. Ther.* 1976;20(6):682-689.
13. Henderson-Smart DJ, Steer P. Methylxanthine treatment for apnea in preterm infants. *Cochrane Database Syst Rev* 2000(2):CD000140.
14. Bhatt-Mehta V, Schumacher RE. Treatment of apnea of prematurity. *Paediatr Drugs* 2003;5(3):195-210.
15. Box GEP, Tiao G. *Bayesian Inference in Statistical Analysis*. New York; 1973.
16. Gelman A, Carlin JB, Stern HS, Rubin DB. *Bayesian Data Analysis*. London: Chapman & Hall; 1995.
17. Bernardo JM, Smith AFM. *Bayesian Theory*. New York: Wiley; 1994.
18. Amzal B, Bois FY, Parent E, Robert CP. Bayesian optimal design via interacting particle systems. *Journal of the American Statistical Association* 2006; 101(474):773-785.
19. Warszawski D, Gorodischer R. Tissue distribution of caffeine in premature infants and in newborn and adult dogs. *Biol. Neonate* 1981;1.
20. International Commission on Radiological Protection (ICRP). *Basic Anatomical and Physiological data for Use in Radiological Protection: Reference Values*. Stockholm, Sweden: Pergamon; 2002.

21. Pons G, Carrier O, Richard M-O, Rey E, D'Athis P, Moran C, et al. Developmental changes of caffeine elimination in infancy. *Dev Pharmacol Ther* 1988;11:258-264.
22. Carrier O, Pons G, Rey E, Richard MO, Moran C, Badoual J, et al. Maturation of caffeine metabolic pathways in infancy. *Clin Pharmacol Ther* 1988;44(2):145-51.
23. Bourgoin H, Paintaud G, Buchler M, Lebranchu Y, Autret-Leca E, Mentre F, et al. Bayesian estimation of cyclosporin exposure for routine therapeutic drug monitoring in kidney transplant patients
Nonparametric estimation of population characteristics of the kinetics of lithium from observational and experimental data: individualization of chronic dosing regimen using a new Bayesian approach. *Br J Clin Pharmacol* 2005;59(1):18-27.
24. Sandstrom M, Karlsson MO, Ljungman P, Hassan Z, Jonsson EN, Nilsson C, et al. Population pharmacokinetic analysis resulting in a tool for dose individualization of busulphan in bone marrow transplantation recipients. *Bone Marrow Transplant* 2001;28(7):657-64.
25. Taright N, Mentre F, Mallet A, Jouvent R. Nonparametric estimation of population characteristics of the kinetics of lithium from observational and experimental data: individualization of chronic dosing regimen using a new Bayesian approach. *Ther Drug Monit* 1994;16(3):258-69.
26. Arnaud M. Metabolism of caffeine and other components of coffee. In: *Coffee and Health*. New York: S Garattini; 1993. p. 43-93.
27. Aranda JV, Collinge JM, Zinman R, Watters G. Maturation of caffeine elimination in infancy. *Arch Dis Child* 1979;54(12):946-9.
28. Le Guennec JC, Billon B, Pare C. Maturation changes of caffeine concentrations and disposition in infancy during maintenance therapy for apnea of prematurity: influence of gestational age, hepatic disease, and breast-feeding. *Pediatrics* 1985;76(5):834-40.
29. Beal SL, Sheiner LB. Estimating population kinetics. *Crit Rev Biomed Eng* 1982;8(3):195-222.
30. Racine-Poon A, Smith AF. Population models. In: Berry DA, editor. *Statistical Methodology in the Pharmaceutical Sciences*. New York: Marcel Dekker, Inc.; 1990. p. 139-162.
31. Garthwaite PH, Kadane JB, O'Hagan A. Statistical methods for eliciting probability distributions. *Journal of the American Statistical Association* 2005;100(470):680-700.
32. Kadane JB, Wolfson LJ. Experiences in elicitation. *Journal of the Royal Statistical Society Series D-the Statistician* 1998;47(1):3-19.
33. Chen MH, Ibrahim JG, Shao QM, Weiss RE. Prior elicitation for model selection and estimation in generalized linear mixed models. *Journal of Statistical Planning and Inference* 2003;111(1-2):57-76.
34. Arnaud M, Weitzholtz H, Voegelin M, Bircher J, Presig R. Assessment of the cytochrome P-448 dependent liver enzyme system by caffeine breath test. In: sato R, editor. *Microsomes drug oxydation and drug toxicity*. New York: Wiley Interscience; 1982. p. 443-444.
35. Stones. Cross-validation choice and assessment of statistical predictions. *Journal of the Royal Statistical Society B* 1974;36:111-147.
36. Gelfand AE, Dey DK, Chang H. Model determination using predictive distributions with implementation via sampling-based methods. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. *Bayesian Statistics 4*. Oxford: Oxford University Press; 1992. p. 147-167.
37. Robert C. *Méthodes de Monte Carlo par Chaînes de Markov*. Paris: Economica; 1996.
38. Gilks WR, Richardson S, Spiegelhalter DJ. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall; 1996.

39. Bois FY, Maszle D. MCSim: a simulation program. *Journal of Statistical Software* 1997;2(9):http://toxi.ineris.fr/activites/toxicologie_quantitative/mcsim/mcsim.php.
40. Chopin N. A sequential particle filter method for static models. *Biometrika* 2002;89:539-552.
41. Tierney L, Mira A. Some adaptive Monte Carlo methods for Bayesian inference. *Statistics in Medicine* 1999;18: 2507-2515.
42. Doucet A, Godsill S, Andrieu C. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing* 2000;10(3):197-208.
43. Doucet A, De Freitas J, Gordon N. *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag; 2001.
44. Ridgeway G, Madigan D. A sequential Monte Carlo method for Bayesian analysis of massive datasets. *Data Mining and Knowledge Discovery* 2003;7(3):301-319.
45. Bayard DS, Jelliffe RW. A Bayesian approach to tracking patients having changing pharmacokinetic parameters. *Journal of Pharmacokinetics and Pharmacodynamics* 2004;31(1):75-107.
46. Geweke J. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* 1989;24:1317-1399.
47. Rubin DB. Using the SIR algorithm to simulate posterior distributions. In: Bernardo JM, De Groot MH, Lindley DV, Smith AFM, editors. *Bayesian Statistics 3*. Oxford: Oxford University Press; 1988. p. 395-402.
48. Tierney L. Markov chains for exploring posterior distributions. *Annals of Statistics* 1994;22(4):1701-1762.
49. Jelliffe R, Schumitzky A, Van Guilder M. Population pharmacokinetics/pharmacodynamics modeling: parametric and nonparametric methods. *Ther Drug Monit* 2000;22(3):354-65.
50. Wakefield J, Walker S. Bayesian nonparametric population models: formulation and comparison with likelihood approaches. *J Pharmacokinet Biopharm* 1997;25(2):235-53.

Tables

Table I: Parameter distributions of our PK model. Summary of the parameter distributions at individual level of the PK model and population prior distributions used in the calibration process.

Parameter (unit)	Distribution	Truncation bounds
Individual level		
α_1 (kg/h)	Normal (μ_{a1}, σ_{a1})	$[-5 \times 10^{-3}; 0]$
α_2 (kg/h)	Normal (μ_{a2}, σ_{a2})	$[0; 3 \times 10^{-3}]$
τ (h)	LogNormal ¹ (μ_τ, σ_τ)	$[72; 480]$
f_{25} (L/kg)	LogNormal ¹ (μ_{f25}, σ_{f25})	$[0.8; 0.9]$
f_τ (L/kg)	LogNormal ¹ ($\mu_{f\tau}, \sigma_{f\tau}$)	$[0.7; 0.8]$
CL_0 (L/h)	LogNormal ¹ (μ_{CL0}, σ_{CL0})	$[0; 0.06]$
γ (L/h ²)	LogNormal ¹ ($\mu_\gamma, \sigma_\gamma$)	$[5 \times 10^{-7}; 10^{-4}]$
cv_M (unitless)	Normal (μ_{cvM}, σ_{cvM})	$[0.010; 0.100]$
Population level		
μ_{a1} (kg/h)	Normal ($-10^{-3}, 0.6 \times 10^{-3}$)	$[-5 \times 10^{-3}; 0]$
μ_{a2} (kg/h)	Normal ($10^{-3}, 0.5 \times 10^{-3}$)	$[0; 3 \times 10^{-3}]$
μ_τ (h)	Uniform	$[72; 480]$
μ_{f25} (L/kg)	Normal (0.85, 0.05)	$[0.8; 0.9]$
$\mu_{f\tau}$ (L/kg)	Normal (0.75, 0.05)	$[0.7; 0.8]$
μ_{CL0} (L/h)	Normal (0.02, 0.01)	$[0; 0.04]$
μ_γ (L/h ²)	Normal ($5 \times 10^{-6}, 5 \times 10^{-6}$)	$[5 \times 10^{-7}; 10^{-4}]$
σ_{a1}	LogNormal ¹ ($0.4 \times 10^{-3}, 2.96$)	-
σ_{a2}	LogNormal ¹ ($0.4 \times 10^{-3}, 2.96$)	-
σ_τ	LogNormal ¹ (1.32, 2.30)	$[1.01; 2]$
σ_{f25}	LogNormal ¹ (1.183, 1.22)	$[1.01; 2]$
$\sigma_{f\tau}$	LogNormal ¹ (1.183, 1.22)	$[1.01; 2]$
σ_{CL0}	LogUniform	$[1.1; 3]$
σ_γ	LogUniform	$[1.1; 3]$
μ_{cvM} (unitless)	Uniform	$[0.010; 0.100]$
σ_{cvM}	Normal (0.030, 0.020)	$[0.005; 0.100]$
σ_C	Normal (1.105, 0.1)	$[1.01; 2]$

¹ For LogNormal distributions, the parameters in parenthesis correspond to the geometric mean (exponential of the mean in log-space) and geometric standard deviation (exponential of the standard deviation in log-space).

Table II: Indicators to compare the existing PK models. Average mean square error (AMSE) and percentages of adequacy data points falling within the ``theoretical" credibility intervals given by various models. These data points concern caffeine concentration of the subjects belonging to the *test set*. 21 data points are available for caffeine concentration in blood.

Model	Nominal credibility interval coverage (%)						AMSE
	40	50	80	85	90	95	
Thomson's <i>et al.</i>	4.8	4.8	33.3	33.3	38.1	38.1	1.358
Lee's <i>et al.</i>	9.5	9.5	14.3	28.6	42.9	47.6	2.207
Falcao's <i>et al.</i>	9.5	14.3	42.9	42.9	47.6	57.1	2.036
New proposed model	48.2	52.4	61.9	71.4	81.0	81.0	0.411

Table III: Summary statistics for the ultimate (N = 35 subjects) marginal posterior distributions of the population parameters.

Population parameters	mean	SD	2.5%tile	97.5%tile
$\mu_{\alpha 1}$	0.00101	0.000174	0.00129	0.000592
$\mu_{\alpha 2}$	0.000948	4.21×10^{-5}	0.000864	0.00103
μ_{τ}	130	15.9	96.3	159
μ_{f25}	0.85	0.0215	0.808	0.892
$\mu_{f\tau}$	0.752	0.0213	0.708	0.793
μ_{CL0}	0.0154	0.00183	0.0121	0.0192
μ_{γ}	8.45×10^{-6}	2.43×10^{-6}	4.02×10^{-6}	1.35×10^{-5}
$\sigma_{\alpha 1}$	0.000676	0.000146	0.000464	0.00103
$\sigma_{\alpha 2}$	0.000237	3.19×10^{-5}	0.000185	0.000309
σ_{τ}	1.6	0.145	1.38	1.93
σ_{f25}	1.21	0.2	1.0	1.69
$\sigma_{f\tau}$	1.21	0.199	1.01	1.7
σ_{CL0}	1.54	0.198	1.2	2
σ_{γ}	1.91	0.456	1.17	2.86
μ_{cvM}	0.0179	0.00462	0.0105	0.0269
σ_{cvM}	0.018	0.00332	0.012	0.025
σ_C	1.36	0.0483	1.28	1.47

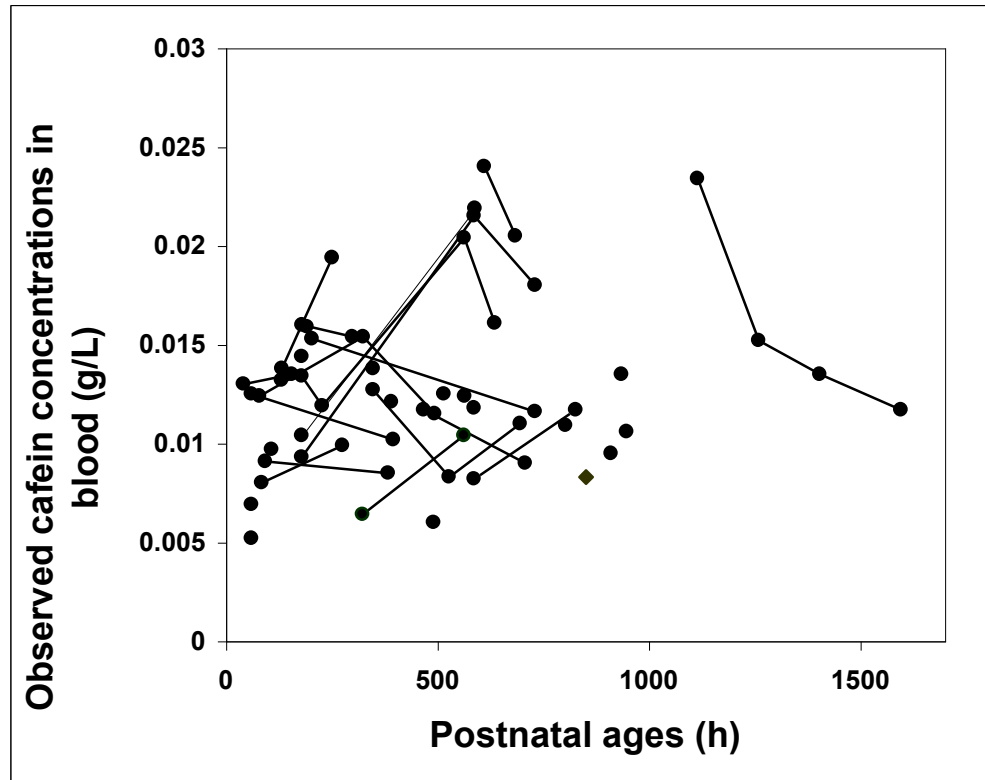


Figure 1: Observed caffeine concentrations in blood as a function of postnatal age. Joined points indicate observations belonging to a same individual.

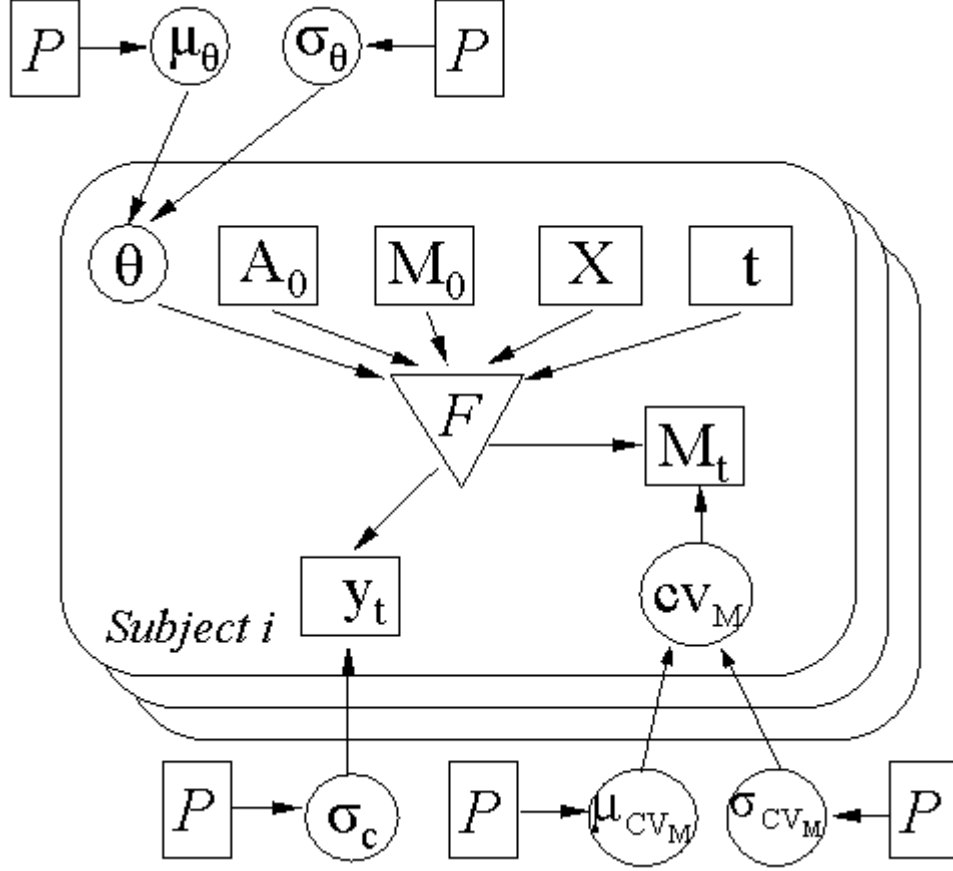


Figure 2: Directed acyclic graph of the hierarchical statistical model. Symbols are: P , prior distributions; μ_θ , population mean parameters; σ_θ , variance of population parameter; t , observation times; θ , unknown parameters; A_0 and M_0 are respectively gestational age and body mass at birth (both supposed exactly measured); X , the administered treatment; M_t , the measured body mass at time t ; cv_M , the variation coefficient of masses and Y_t , the measured caffeine concentration in blood. μ_{cv_M} , and σ_{cv_M} are respectively for mean and variance of cv_M parameters and σ_c for the variance of the experimental measurements. F is the dynamic bio-distribution model. Square nodes are for variables of known (or supposed known) values; circular nodes for unknown variables and the triangle represents a deterministic link.

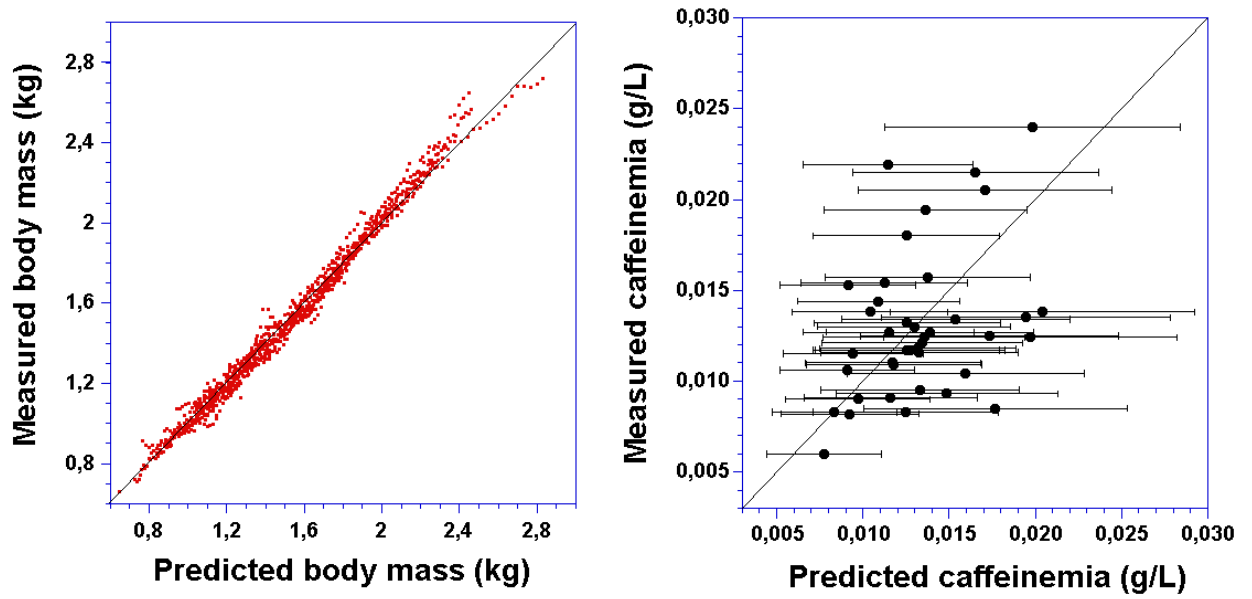


Figure 3: Data versus model predictions. The left and the right graphs respectively concern body masses and caffeine concentrations in blood for individuals of the *training* set. On the right side graph, each line segment describes the 90%-credibility interval associated to the corresponding prediction. The 90%-credibility intervals have been plotted using σ_c value of the maximum posterior vector.

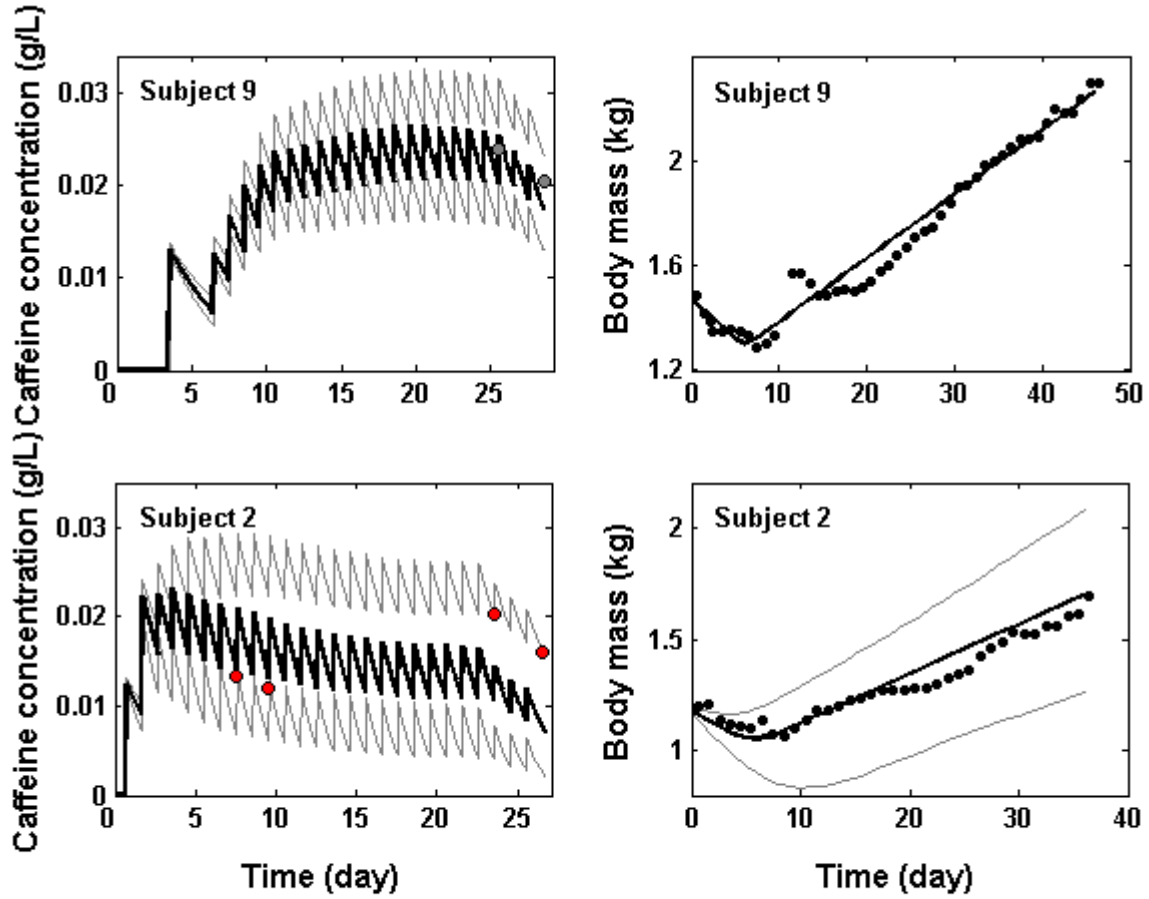


Figure 4: Caffeine concentration in blood (left) and body masses (right) as a function of time. Predictions have been calculated for one subject of the *training set* (Subject 9) and one subject of the *test set* (Subject 2). Means of the predictions are in bold lines, and the corresponding 90%-credibility intervals are plotted with thin lines. Data are represented by dots. Note that 90%-credibility interval of the predicted masses is not represented for the subject of the calibration set (subject 9) as it was very thin.

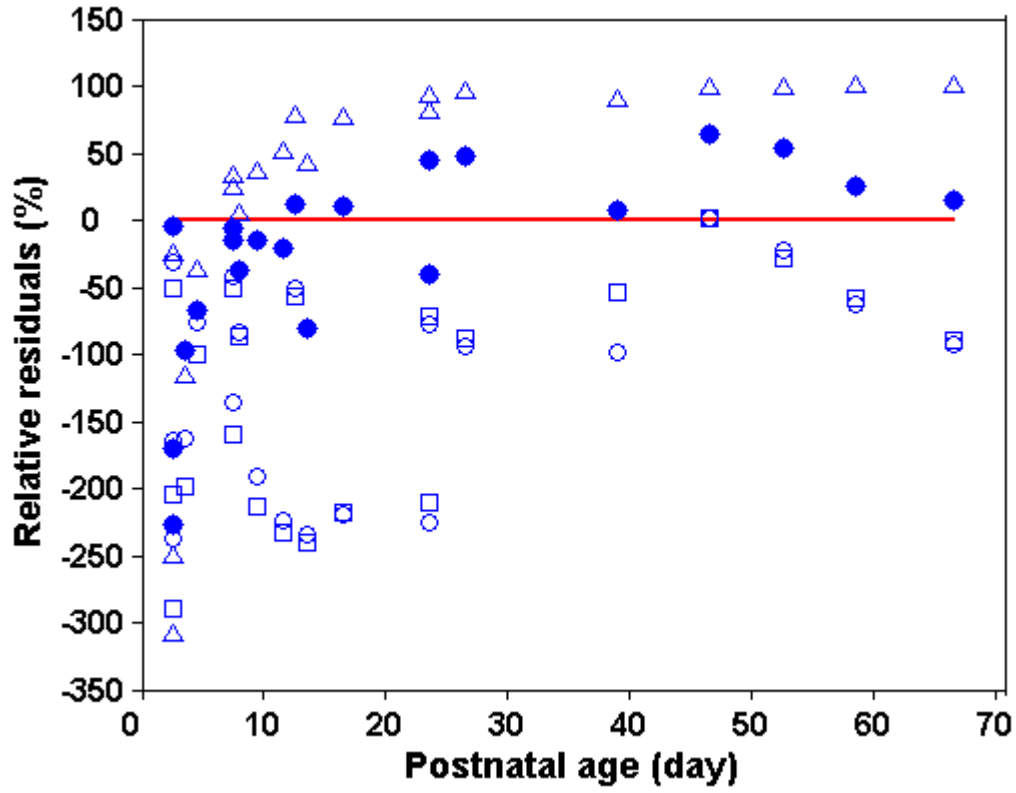


Figure 5: Relative residuals for caffeine concentration as a function of postnatal age. Four different models are represented: Thomson *et. al.* ^[1] (triangles), Lee *et. al.* ^[2] (squares), Falcao *et. al.* ^[3] (empty circles) and the one we propose (plain circles). Relative residuals correspond to the relative differences between *training set* data points and mean caffeine concentration predictions obtained by Monte Carlo simulations.

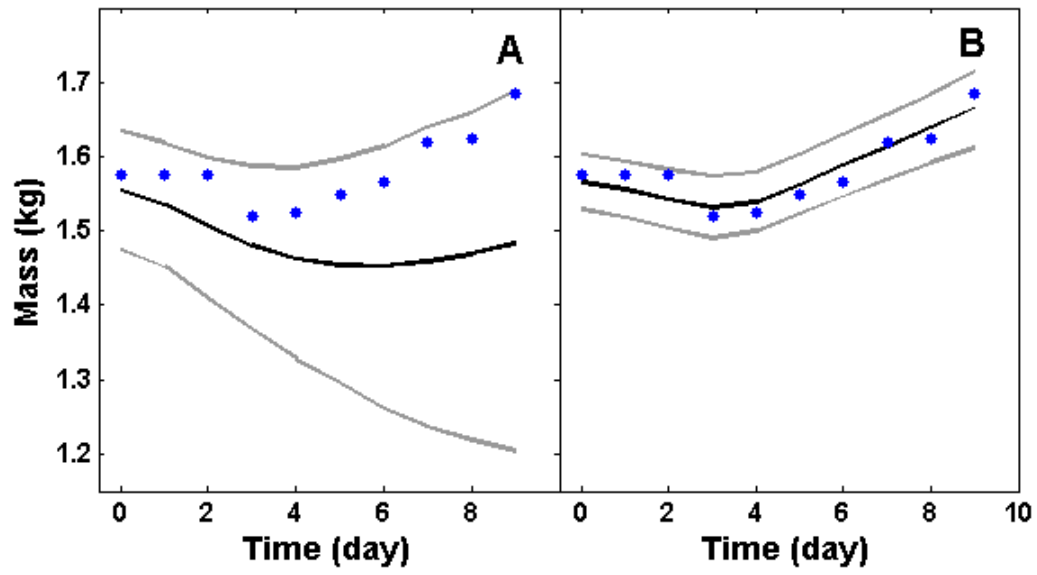


Figure 6: Body mass predictions as a function of the time for patient 31. The 5 %, and 95 % quantiles of predictions are plotted using gray lines, the mean of predictions using dark bold lines. Data are represented by the dots. Panel A shows the predicted masses calculated from the original posterior-sampled parameters, using the covariates of the patient at birth, only. Predictions of panel B use, in addition, body mass data via the updated posterior parameters obtained by a particle algorithm.

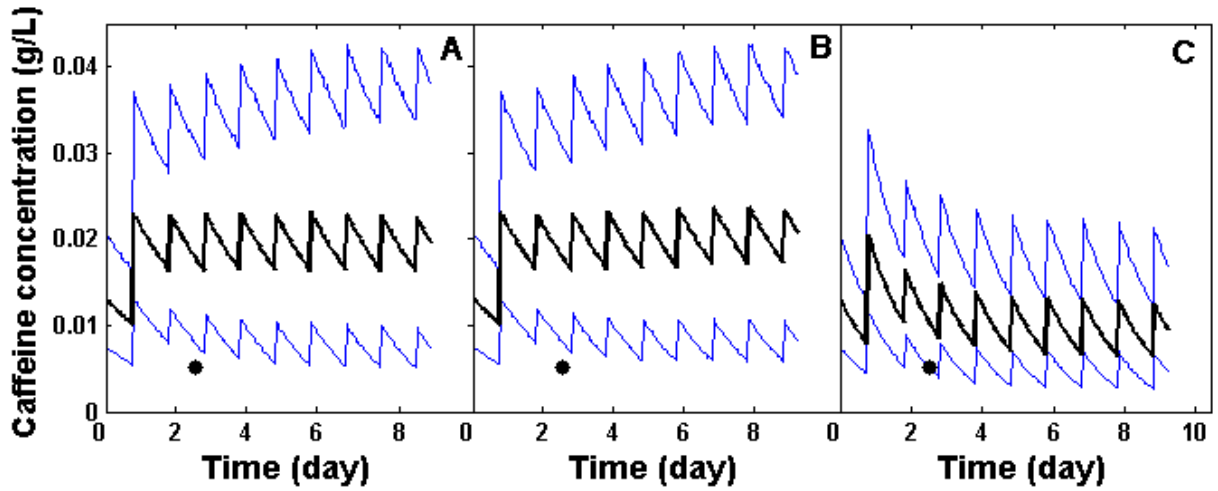


Figure 7: Caffeine concentration predictions as a function of time for patient 31. The 5 % and 95 % quantiles of predictions are plotted using thin lines, the mean of predictions using bold lines. The data point is represented by dot. Panel A shows the predicted caffeine concentration in blood calculated from original posterior sampled parameters and using only the covariates of the patient. Predictions of panel B and C are performed using the posterior parameters updated *via* a particle algorithm. In panel B, only data masses were used for parameter updating while only measured caffeine concentration was used to update predictions in panel C.

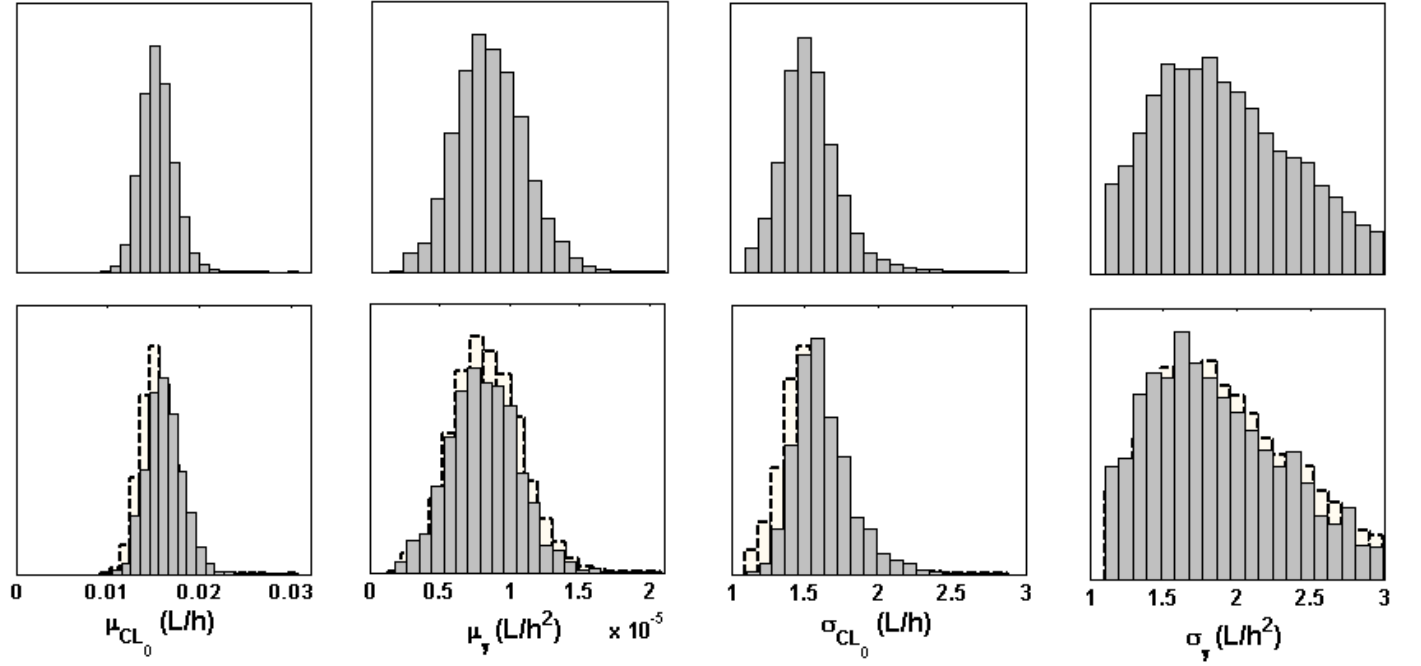


Figure 8: Histograms of the marginal posterior distributions of the population elimination model parameters. Top row graphs give the marginal distributions of the *original* posterior. Bottom row darker histograms give the marginal updated distributions, using one caffeine concentration data point for subject 31. Bottom row unfilled histograms recall the original posterior.